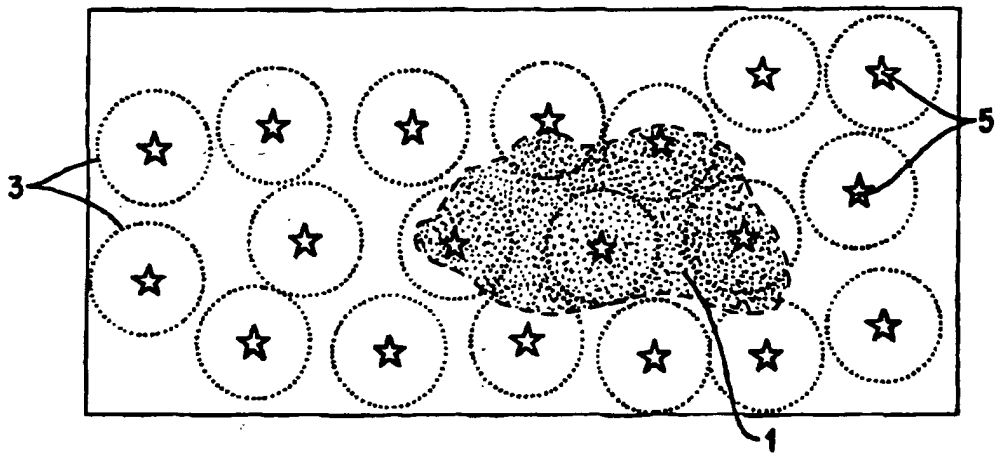## PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06F 19/00	A1	(11) International Publication Number: **WO 97/27559**
		(43) International Publication Date: 31 July 1997 (31.07.97)

<table>
<tr><td>

(21) International Application Number: PCT/US97/01491

(22) International Filing Date: 27 January 1997 (27.01.97)

(30) Priority Data:
08/592,132   26 January 1996 (26.01.96)   US
08/657,147   3 June 1996 (03.06.96)   US

(71)(72) Applicants and Inventors: PATTERSON, David, E. [US/US]; 1908 Bookbinder Drive, St. Louis, MO 63146 (US). CRAMER, Richard, D. [US/US]; 9012 Highway DD, O'Fallon, MO 63366 (US). CLARK, Robert, D. [US/US]; 827 Renee Lane, St. Louis, MO 63141 (US). FERGUSON, Allan, M. [GB/US]; 2314 Callender Court, St. Louis, MO 63017 (US).

(74) Agent: WEINBERGER, Laurence, A.; Suite 103, 882 S. Matlack Street, West Chester, PA 19382 (US).

</td><td>

(81) Designated States: AU, CA, CN, CZ, HU, IL, JP, KR, NO, PL, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

</td></tr>
</table>

(54) Title: METHOD OF CREATING AND SEARCHING A MOLECULAR VIRTUAL LIBRARY USING VALIDATED MOLECULAR STRUCTURE DESCRIPTORS

(57) Abstract

The problem of how to select out of a large chemically accessible universe molecules representative of the diversity of that universe is resolved by the discovery of a method to validate molecular structural descriptors. Using the validated descriptors, optimally diverse subsets (5) can be selected. In addition, from the universe, molecules with characteristics similar to a selected molecule can be identified (3). The validated descriptors also enable the generation of a huge virtual library of potential product molecules which could be formed by combinatorial arrangement of structural variations and cores. In this virtual library it is possible to search billions of possible product compounds in relatively short time frames.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/01491

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G06F 19/00
US CL :364/496

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 364/496-499; 395/601,616

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,307,287 A (CRAMER, III ET AL) 26 April 1994, see abstract.	52-54
X	US 5,025,388 A (CRAMER, III ET AL) 18 June 1991, see abstract.	52-54
A	US 5,345,516 A (BOYER ET AL) 09 September 1994, see entire document.	1-94
A	US 5,270,170 A (SCHATZ ET AL) 14 December 1993, see entire document.	1-94

☐ Further documents are listed in the continuation of Box C.   ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"E"	earlier document published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 APRIL 1997	2 8 MAY 1997

Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer EMANUEL T. VOELTZ
Facsimile No. (703) 305-3230	Telephone No. (703) 305-9714

Form PCT/ISA/210 (second sheet)(July 1992)*

03/05/2003, EAST Version: 1.03.0002

if no activity is detected, stop; and

h. Repeating steps e through g until no further compounds show activity in the assay.

52.   A computer-based method of characterizing the three dimensional structure of reactants, which can assume many conformations, comprising the steps of:

5   a. Topomerically aligning the reactants; and

    b. Determining the CoMFA steric fields for each topomerically aligned reactant.

53.   The method of claim 52 further comprising the addition of topomeric hydrogen bonding fields to the CoMFA steric fields.

54.   A computer-based method of applying a molecular structural descriptor to a set of

10 reactants comprising the following steps:

    a. Topomerically aligning the reactants;

    b. Determining the CoMFA steric fields for each topomerically aligned reactant; and

    c. Calculating the field differences between all pairs of reactants.

55.   The method of claim 54 further comprising after step b the additional step of adding

15 topomeric hydrogen bonding fields to the CoMFA fields.

56.   The method of claim 54 further comprising after step c the additional step of hierarchically clustering the reactants until the intercluster distance is about 80 - 100 CoMFA field units.

57.   In a digital computer in which representations of specified reactant molecules and a

20 core molecule have been stored, a computer-based method for selecting, for all possible product molecules which could be created in a combinatorial synthesis from the reactant molecules and common core molecule, a subset of product molecules, comprising the following steps:

    a. Characterizing all the reactant molecules with a validated molecular structural

25   descriptor appropriate to reactant molecules;

    b. Hierarchically clustering the characterized reactant molecules until the intercluster
        distance corresponds to the neighborhood distance of the validated molecular
        structural descriptor or to a value close to the neighborhood distance which creates
        a logical clustering break;

30   c. Selecting a reactant molecule from each cluster;

    d. Combinatorially assembling the selected reactant molecules and core molecule into
        products which would be created in the chemical synthesis;

    e. Selecting a product molecule for inclusion in the subset;